On the unification of families of skew-normal distributions

Reinaldo B. Arellano-Valle
Departamento de Estadística
Pontificia Universidad Católica de Chile
Santiago, Chile

Adelchi Azzalini
Dipartimento di Scienze Statistiche
Università di Padova
Italia

28th October 2004

Abstract

The distribution theory literature connected to the multivariate skew-normal distribution has grown rapidly in recent years, and a number of extensions and alternative formulations have been put forward. Theretofore, at the moment there is a variety of coexisting proposals, similar but not identical, and with rather unclear connections. The purpose of this paper is to unify the various proposals under a new general formulation, clarifying at the same time the relationships with and among other proposal. The final part sketches an extension of the argument to the skew-elliptical family.

Some key-words: skew-normal distribution, skew-elliptical family, skew-t distribution, stochastic representation.

1 The skew-normal distribution and its descendants

1.1 General remarks

There has been recently a resumption of interest in the study of parametric classes of probability distributions for continuous multivariate random variables. A substantial fraction of this activity has stemmed by the proposal on the multivariate skew-normal (SN) distribution, which represents an extension of the multivariate normal family to which an additional parameter is being added to regulate skewness. The SN class has been studied by Azzalini & Dalla Valle (1996) and by Azzalini & Capitanio (1999), and subsequently by several authors, with intense developments in various directions, often considering non-normal symmetric families as the starting point in place of the normal class.

These further developments are now so numerous that is not feasible to recall here all of them; hence we shall restrict our references to specific contributions, as their relevance occurs in the exposition of the paper. For reviews of the research work produced in this area, see Arnold & Beaver (2002) and Azzalini (2003). Here we shall confine ourselves to a single remark: while the original motivation of this approach was the introduction of skewness in the normal family of distributions, the current developments are of a much higher level of generality and the effect of these extensions to standard parametric families is much more substantial than simple 'insertion' of skewness.

This very active and stimulating context has seen the appearance of several variants or alternative proposals of the original skew-normal distribution. Among these, we mention in particular: the Closed Skew-Normal (CSN) of González-Farías et al. (2004), the Hierarchical Skew-Normal (HSN) of Liseo & Loperfido (2003), the Fundamental Skew-Normal (FSN) of Arellano-Valle & Genton (2005), the Multivariate Skew-Normal (MSN) of Gupta et al. (2004), the Skew-Normal of Sahu et al. (2003); although on a slightly different direction, see also Arnold & Beaver (2000)

While the existence of some many proposals is a sign of the high vitality of this stream of literature, it inevitably poses problems, namely: is there one version which is the 'overall best' skew-normal distribution? are some of these proposals equivalent up to a reparametrisation? more generally, what are the connections among them? Answering these questions is relevant for various reasons, beside the mathematical interest *per se*. First of all, the co-existence of several alternative variants is a source of much difficulty and confusion, particularly for a non-specialist who wants to approach this area. In addition, each group of authors tend naturally to develop further their own formulation, and the final effect would eventually be a lack of communication among what so far has been a closely interlinked branch of literature.

The purpose of the present paper is to tackle the above questions, and to re-constitute a 'unified formulation' of the skew-normal distribution. To achieve this target we shall introduce a further type of skew-normal distribution, and this very fact seems to contradict our purpose. We show however that this new formulation encompasses all previous proposals, once some redundancies in parametrisation are removed, and this process also clarifies the connections among previous variants. In addition, reasons are given for preferring certain forms of parametrisation in place of others.

1.2 The basic skew-normal distribution

To establish notation, the density function at point x ($x \in \mathbb{R}^d$) of a $N_d(\mu, \Sigma)$ random variable will be denoted by $\phi_d(x-\mu; \Sigma)$; similarly, $\Phi_d(x-\mu; \Sigma)$ represents the corresponding distribution function. When $d = 1$, we omit the subscript of ϕ and Φ.

The original version of the SN distribution refers to the d-dimensional density function whose value at x ($x \in \mathbb{R}^d$) is given by

$$2\,\phi_d(x - \xi; \Omega)\,\Phi\left(\alpha^\top \omega^{-1}(x - \xi)\right) \tag{1}$$

where ξ ($\xi \in \mathbb{R}^d$) is a location parameter, Ω is a positive definite covariance matrix, α ($\alpha \in \mathbb{R}^d$) is a parameter which regulates skewness, and ω is a diagonal matrix formed by the standard deviations of Ω; hence $\Omega = \omega\bar{\Omega}\omega$, where $\bar{\Omega}$ is a correlation matrix. If a random variable $Y = (Y_1, \ldots, Y_d)^\top$ has density (1), we write $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$.

Clearly, setting $\alpha = 0$ in (1) produces the $N_d(\xi, \Omega)$ distribution. Besides this simple fact, many other formal properties and analogies with the normal family of distributions support the adoption of the term 'skew-normal', but we do not review these aspects here, and refer the reader to the literature quoted above. What is more important for our development are two forms of stochastic representation for a random variable of SN type.

The first of these representations is obtained via the following conditioning mechanism. Denote by U_0 and U_1 two random variables of dimension 1 and d, respectively, such that

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{1+d}(0, \Omega^*), \qquad \Omega^* = \begin{pmatrix} 1 & \delta^\top \\ \delta & \bar\Omega \end{pmatrix} \tag{2}$$

where $\bar\Omega$ is a correlation matrix and

$$\delta = (1 + \alpha^\top \bar\Omega \alpha)^{-1/2} \bar\Omega \alpha, \tag{3}$$

which ensures that Ω^* is a correlation matrix. Then $Z = (U_1 | U_0 > 0)$ has density function

$$2\,\phi_d(x; \bar\Omega)\;\Phi(\alpha^\top x) \tag{4}$$

and the affine transformation

$$Y = \xi + \omega Z$$

has density function (1). For later use, we recall a form of dual expression of (3), given by

$$\alpha = (1 - \delta^\top \bar\Omega^{-1} \delta)^{-1/2} \bar\Omega^{-1} \delta\,. \tag{5}$$

The other stochastic representation is of convolution type. Specifically, assume that V_0 and V_1 are independent variables with distribution $N(0,1)$ and $N_d(0, \Psi)$, respectively, where Ψ is a correlation matrix; also, let $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_d)$ where $\delta_j \in (-1, 1)$ for all j's, denote by I_d the identity matrix of order d and by 1_d the d-dimensional vector of all 1's. Then

$$Z = \Delta\, 1_d\, |V_0| + (I_d - \Delta^2)^{1/2}\, V_1 \tag{6}$$

has distribution of type (4), with a known relationship between the (Ψ, Δ) and the $(\bar\Omega, \alpha)$ sets of parameters. Again, the affine transformation $Y = \xi + \omega Z$ leads to density (1).

1.3 Some extensions and variants

The basic form (1) lends itself to a number of extensions, which have been extensively studied in the papers quoted in § 1.1 and others publications mentioned therein. These extensions follow one, or possibly more, of the following directions of work.

1. Replace the 0 value of $\mathbb{E}\{U_0\}$ in (2) by a further parameter, γ say. This additional parameter changes the normalising constant in (1) from 2 to $1/\Phi(\gamma)$, and it prevents the stochastic representation (6), at least in the exact form given above. Notice that another variant is equivalent, namely setting $Z = (U_1 | U_0 > \tau)$, but the two parameters τ and γ cannot be included simultaneously, since only $\tau - \gamma$ is identifiable.

   A stochastic representation similar to (6) is given by

   $$Z = \Delta\, 1_d\, V_0(-\gamma) + (I_d - \Delta^2)^{1/2}\, V_1$$

   where $V_0(c) \sim LTN(c; 0, 1)$, which represents the distribution obtained by truncation below c of a $N(0,1)$ variate. Notice that the representation $V_0(c) = (V_0 | V_0 > c)$, with $V_0 \sim N(0,1)$ holds.

   This form of representation has been obtained by Arnold & Beaver (2002). It is also connected with a discussion, started by Weinstein (1964) and summarised by Nelson (1964), which contains in an implicit form the ingredients leading to we now call skew-normal distribution.

2. The single hidden variable U_0 in (2) can be replaced by a multivariate variable. Correspondingly, there is a set of constraints of type $U_0 > 0$ where it is understood that the inequality must hold for each of the components of U_0. A similar form of multivariate extension has been considered for the convolution representation (6). The various formulations of this type will be discussed more in detail in § 2.2 below.

3. Much work has been done replacing the normality assumption by a more general one, especially using densities of the elliptical families. In this paper, an extension of our main argument to the elliptical family is sketched in the final section, but we do not dwell into the connections with other proposals.

Some of the variants which have been put forward occur as reparametrisations of the above formulation, and some comments are in order.

It is algebraically simpler and apparently equivalent to drop the term ω inside the argument of $\Phi(\cdot)$ in (1), but in doing so α looses its interpretation as 'skewness parameter'. To see this, suppose that $Y \sim \mathrm{SN}_d(0, \Omega, \alpha)$, where we take the location parameter to be 0 for simplicity of argument, and consider the scaled version $\tilde{Y} = \tilde{\omega} Y$, for some diagonal matrix $\tilde{\omega}$. In the formulation given above, $\tilde{Y} \sim \mathrm{SN}_d(0,\, \tilde{\omega}\,\Omega\,\tilde{\omega},\, \alpha)$ with α unchanged, but this is not the case in the variant parametrisation, since its α parameter changes to $\tilde{\omega}^{-1}\alpha$. It is clearly unsatisfactory that a skewness parameter changes its value as an effect of change of scale.

It is true that even α does not reflect only the skewness of the components, since it depends on the shape of the marginals as well as the correlation matrix $\bar{\Omega}$, although at least it does not depend on the scale parameter ω. In this sense, the use of δ in place of α is preferable, since its generic component δ_j is directly related to the index of skewness of the corresponding component Y_j. There is however the drawback that δ and $\bar{\Omega}$ are not variation-independent, as α and $\bar{\Omega}$ are.

Another variant form of (1) starts replacing Ω in (2) with the identity matrix, and set $Z = (U_1 | U_0 > 0)$. To incorporate a covariance parameter, define further $Y' = \xi + \Omega^{1/2}\, Z$ for some square root $\Omega^{1/2}$ of a covariance matrix Ω. The resulting density involves then a term $\Omega^{-1/2}$ which is problematic when we come to state a 'definition' of skew-normal distribution, since the square root is not unique, except when Ω is diagonal. Consequently, one must either suffer the fact that the 'definition' effectively defines several distributions, or to impose somewhat arbitrarily a specific choice of square root, such as the Cholesky decomposition. The latter option has also the disadvantage of becoming hardly acceptable when the skewness parameter is 0, since it would be equivalent to parameterise the normal family by $(\xi, \Omega^{1/2})$ instead of (ξ, Ω).

2 Yet another skew-normal distribution

2.1 Two equivalent representations

We introduce an extension of the basic SN distribution via two equivalent types of constructions. The first of these is based on a representation of type (2), but with U_0 of dimension m, and the new constraint $U_0 + \gamma > 0$ for some $\gamma \in \mathbb{R}^m$. As already indicated, a notation of type $X > 0$, when X is a vector, means that each component of X exceeds 0.

Assume that (U_0, U_1) is jointly a multivariate normal variable of dimension $m + d$ with distribution

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{m+d}\left(0, \Omega^*\right), \qquad \Omega^* = \begin{pmatrix} \Gamma & \Delta^\top \\ \Delta & \bar\Omega \end{pmatrix} \tag{7}$$

where Ω^* is a positive definite correlation matrix, and consider the distribution of $Z = (U_1 | U_0 + \gamma > 0)$.

The density function of a variable of type $(X_1 | X_0 > c)$, evaluated at point x, is most easily computed via the general relationship

$$f(x) = \frac{f_{X_1}(x)\, \mathbb{P}\{X_0 > c | X_1 = x\}}{\mathbb{P}\{X_0 > c\}}, \tag{8}$$

in an obvious notation. The role of this simple yet important relationship, clearly valid also outside the normal context, has been stressed by Arellano-Valle et al. (2002); see their Theorem 5.1.

The case under consideration lends itself to simple calculation, since $(U_0 + \gamma | U_1 = y)$ has still a distribution of normal type, with well-know expression of the parameters. After simple algebra, one obtains that the density function of $Y = \xi + \omega Z$ is

$$f(y) = \phi_d(y - \xi; \Omega)\, \frac{\Phi_m(\gamma + \Delta^\top \bar\Omega^{-1} \omega^{-1}(y - \xi); \Gamma - \Delta^\top \bar\Omega^{-1} \Delta)}{\Phi_m(\gamma; \Gamma)} \tag{9}$$

for $y \in \mathbb{R}^d$. For reasons explained in § 2.2, we shall call this expression the Unified Skew-Normal density and, to ease pronunciation, we adopt the acronym SUN. Hence we write $Y \sim \mathrm{SUN}_{d,m}(\xi, \gamma, \bar\omega, \Omega^*)$, where $\bar\omega$ is the vector of the diagonal elements of ω, i. e. $\bar\omega = \omega\, 1_d$, and it is intended that the terms $\Gamma, \Delta, \bar\Omega$ required by (9) are recovered by appropriate partitioning of Ω^* as indicated in (7).

There exists another form of genesis for distribution (9), using a convolution mechanism instead of conditioning. Let V_0 and V_1 be independent variables such that

$$V_0(-\gamma) \sim LTN_m(-\gamma; 0, \Gamma), \qquad V_1 \sim N_d(0, \Psi) \tag{10}$$

where Γ and Ψ are correlation matrices, and the notation $LTN_m(c; \mu, \Sigma)$ denotes a multivariate normal variable with all components truncated below c, and consider the transformation

$$Y' = \xi + \omega\, \{B_0\, V_0(-\gamma) + B_1\, V_1\} \tag{11}$$

where B_0 and B_1 denote non-random matrices of order $d \times m$ and $d \times d$ matrix, respectively, such that

$$\bar\Omega = B_0 \Gamma B_0^\top + B_1 \Psi B_1^\top, \quad \Delta = B_0 \Gamma. \tag{12}$$

The matrix B_1 can be chosen arbitrarily, under the above constraints. To see that the set of possible choices is not empty, consider $B_1 = (I_d - \bar\Delta^2)^{1/2}$, where $\bar\Delta^2$ is the diagonal matrix whose diagonal elements are the same as the diagonal elements of $\Delta \Gamma^{-1} \Delta^\top$. In particular, with this choice of B_1, (11) reduces to (6) when $m = 1$, $\gamma = 0$ and $\Gamma = 1$.

Taking into account that the moment generating function of $V_0(-\gamma)$ evaluated at s is

$$\exp(\tfrac{1}{2} s^\top \Gamma s) \frac{\Phi_m(\gamma + \Gamma s; \Gamma)}{\Phi_m(\gamma; \Gamma)}, \qquad (s \in \mathbb{R}^m)\,,$$

it follows that, under conditions (12), the moment generating function of Y' is

$$M(t) = \exp(\xi^\top t + \tfrac{1}{2} t^\top \Omega t) \, \frac{\Phi_m(\gamma + \Delta^\top \omega t; \Gamma)}{\Phi_m(\gamma; \Gamma)}, \qquad (t \in \mathbb{R}^d) \tag{13}$$

and the corresponding density function is (9); hence Y and Y' are equal in distribution.

Representation (11) is the analogue of (6), with the main variant that the term $|V_0|$ of the latter is replaced by the $V_0(-\gamma)$. The reason for the change is that taking the absolute value of the components of $V_0 \sim N_m(\gamma, \Gamma)$ does not produce any meaningful distribution. For $\gamma = 0$, an exception to this is the special case where Γ is proportional to I_m, a situation which will be discussed separately in Section 2.2, in connection with the FSN family.

One limitation of the basic SN distribution (1) is that, if we partition Z into two components, Z_1 and Z_2, say, then it allows independence between Z_1 and Z_2 only if one of the two has a symmetric normal distribution. This limitation is overcome by the SUN family, which removes that condition for independence; the required conditions on the structure of the matrix Ω^* in (7) are given in an appendix. On applying these conditions recursively to the individual subvectors Z_1 and Z_2, it is seen that at most m independent components are possible.

2.2 Relationships with similar families

The aim of this section is to show that the SUN family encompasses all extensions to the basic SN family mentioned in § 1.1, provided redundancies in the parametrisation of some proposals are removed.

Basic SN family and a simple extension It is easy to see that the same of the basic $\mathrm{SN}(\xi, \Omega, \alpha)$ distribution is the same of the $\mathrm{SUN}_{d,1}(\xi, 0, \omega, \Omega^*)$ distribution, except for a different form of representation; here $\Gamma = 1$ and Δ is given by (3). Conversely, given that a random variable has distribution $\mathrm{SUN}_{d,1}(\xi, 0, \omega, \Omega^*)$, the corresponding $\mathrm{SN}_d(\xi, \Omega, \alpha)$ distribution is uniquely identified, taking into account (5).

A simple extension of the above case is given by $\mathrm{SUN}_{d,1}(\xi, \gamma, \omega, \Omega^*)$ which corresponds to the distribution examined by Arnold & Beaver (2000) and by Capitanio et al. (2003), again up to change of parametrisation.

Closed SN family González-Farías et al. (2004) have examined a form of representation via conditioning similar to our, and obtained a class of distributions called the Closed Skew-Normal (CSN) family. The main difference from our conditioning mechanism is that the analogue of the covariance matrix appearing in (7) is of form

$$\begin{pmatrix} \Theta + D\Sigma D^\top & D\Sigma \\ \Sigma D^\top & \Sigma \end{pmatrix}$$

whose ingredients are not required to satisfy any condition, except that $\Sigma > 0$ and $\Theta > 0$. The CSN family is then obtained as the distribution of $Y = \xi + Z$, where $Z = (U_1 | U_0 + \gamma > 0)$, and parameterised as $CSN_{d,m}(\xi, \Sigma, D, \gamma, \Theta)$, except that we have introduced a slight change of notation.

However, the lack of scale constraints on the above covariance matrix makes the parametric class not identifiable. To see this, consider a diagonal matrix G which can be chosen arbitrarily provided its diagonal terms are all positive, and notice that the condition $U_0 + \gamma > 0$ is equivalent to $GU_0 + G\gamma > 0$. Therefore $CSN_{d,m}(\xi, \Sigma, D, \gamma, \Theta)$ and $CSN_{d,m}(\xi, \Sigma, GD, G\gamma, G\Theta G)$ refer to the same probability distribution.

Another critical aspect of the adopted parametrisation is D, which is described as a skewness parameter. In fact, D is not invariant to changes of scale, for reasons analogous to those discussed in § 1.3.

Once the above aspects of the parametrisation are adjusted, the SUN and the CSN class are equivalent, on setting $\bar{\Omega} = \Sigma$, $\Gamma = \Theta + D\Sigma D^\top$ and $\Delta = \Sigma D^\top$.

CSN-2 family Another extension of the basic SN family is given by the density

$$f(y) = \phi_d(y - \xi; \Psi + D^\top \Gamma D) \frac{\Phi_m(\gamma + \Gamma D(\Psi + D^\top \Gamma D)^{-1}(y - \xi); (\Gamma^{-1} + D\Psi^{-1}D^\top)^{-1})}{\Phi_m(\gamma; \Gamma)} \quad (14)$$

which is an extension a distribution considered by Arellano-Valle & Genton (2005); when $\xi = 0$, $\gamma = 0$ and $\Gamma = I_m$, this density reduces to their (1.9). That density is in turn an extension of the family introduced by Sahu et al. (2003).

Since it can be shown that, similarly to the earlier CSN family, also the family of densities (14) is closed by marginalisation and conditioning, then we refer to is as the CSN-2 class. If a random variable Y has density function (14), we write $Y \sim \mathrm{CSNII}_{d,m}(\xi, \gamma, \Psi, \Gamma, D)$.

The above density function can be obtained by a conditioning mechanism, similar to those discussed earlier, on setting

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{m+d}(0, \Omega^{**}), \quad \Omega^{**} = \begin{pmatrix} \Gamma & \Gamma D \\ D^\top \Gamma & \Psi + D^\top \Gamma D \end{pmatrix}$$

and considering $Y = \xi + Z$, where $Z = (U_1 | U_0 + \gamma > 0)$. An alternative construction of this distribution will be discussed later. A peculiar aspect of this density is that the skewness matrix D enters into the density ϕ_d which represents the 'symmetric part' of the density.

Even for this class there is a problem of overparametrisation, if the covariances matrices are not restricted. In fact $\mathrm{CSNII}_{d,m}(\xi, G\gamma, \Psi, G\Gamma, G^{-1}D)$, which correspond to distributions of $(U_1 | GU_0 + G\gamma > 0)$, for any diagonal matrix $G > 0$, and $\mathrm{CSNII}_{d,m}(\xi, G\gamma, \Psi, G\Gamma, G^{-1}D)$, correspond to the same density. The coincidence with the SUN family is achieved on setting $\bar{\Omega} = \Psi + D\Gamma D^\top$ and $\Delta = D^\top \Gamma$.

Hierarchical SN family The Hierarchical Skew-normal (HSN) distribution has been obtained by Liseo & Loperfido (2003) within a Bayesian context, but the purely probabilistic argument can linked to our construction as follows. If $\theta_0 \sim N_d(0, \Sigma)$ and $\theta_1 \sim N_d(0, \Upsilon)$ are independent variables, and C is a full-rank $m \times d$ matrix ($1 \leq m \leq d$), set

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} = \begin{pmatrix} -C\theta_0 \\ \theta_0 + \theta_1 \end{pmatrix} \sim N_{m+d}\left(0, \begin{pmatrix} C\Upsilon C^\top & -C\Upsilon \\ -\Upsilon C^\top & \Upsilon + \Sigma \end{pmatrix}\right)$$

and define the HSN distribution is as the one of the variable $\mu + (U_1 | U_0 > c)$, parameterised as $HSN_d(\mu, c, \Sigma, \Upsilon, C)$.

There is however a problem of over-parametrisation, very similar to the one of the CSN family, since even here it originates by the lack of suitable scale constraints. The condition $U_0 > c$ is the same of $GU_0 > Gd$ if G is as above. Hence the notations $HSN_d(\mu, c, \Sigma, \Upsilon, C)$ and $HSN_d(\mu, Gc, \Sigma, \Upsilon, GC)$ refer to the same distribution.

Fundamental SN family The Fundamental Skew-normal (FSN) family studied by Arellano-Valle & Genton (2005) is obtained via a convolution mechanism which generalises (6), rather than of conditioning type. An analogous representation of conditioning type is however not difficult to construct. In it canonical form, $\mathrm{CFSN}_{d,m}(\Delta)$ say, the distribution coincides with $\mathrm{SUN}_{d,m}(0, 0, I_d, \Omega^*)$, where

$$\Omega^* = \begin{pmatrix} I_m & \Delta^\top \\ \Delta & I_d \end{pmatrix}.$$

This form does not include location and scale parameter, and a dependence structure separate from the one induced by the skewness parameter Δ. To insert these ingredients starting from a variable Z with distribution of above type, one sets $Y = \mu + \Sigma^{1/2} Z$, for some positive definite matrix $\Sigma^{1/2}$. This choice involves the use of a square root matrix $\Sigma^{1/2}$, whose critical aspects have been already discussed in § 1.3.

The next table summarises the correspondence between the parametrisation of the SUN and those of the other families, assuming all scale factors are 1's and location parameters are 0.

SUN	$\bar{\Omega}$	Γ	Δ
CSN-1	Σ	$\Theta + D\Sigma D^\top$	$(D\Sigma)^\top$
CSN-2	$\Psi + D^\top \Gamma D$	Γ	$(\Gamma D)^\top$
HSN	$\Upsilon + \Sigma$	$C\Upsilon C^\top$	$-(C\Upsilon)^\top$
CFSN	I_d	I_m	Δ

2.3 Correspondence between the two representations of the SUN

In this section, we want to examine the connection between the two stochastic representations of the SUN family set out in Section 2.1. To this end, consider two $(m + d)$-dimensional Normal random variables, U and V, having standardised marginals and

$$\mathrm{var}\{U\} = \mathrm{var}\left\{\begin{pmatrix} U_0 \\ U_1 \end{pmatrix}\right\} = \begin{pmatrix} \Gamma & \Delta^\top \\ \Delta & \bar{\Omega} \end{pmatrix}, \qquad \mathrm{var}\{V\} = \mathrm{var}\left\{\begin{pmatrix} V_0 \\ V_1 \end{pmatrix}\right\} = \begin{pmatrix} \Gamma & 0 \\ 0 & \Psi \end{pmatrix},$$

where the U_0 and the V_0 blocks are m-dimensional. The connection between these sets of variables can be examined on writing

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} = \begin{pmatrix} A_0 V_0 + A_1 V_1 \\ B_0 V_0 + B_1 V_1 \end{pmatrix} \tag{15}$$

and considering the implications for the non-stochastic matrices (A_0, A_1, B_0, B_1).

There are indeed many choices of these matrices to achieve the above distributional equality. One particularly meaningful option is to set

$$V_0 = U_0, \qquad V_1 = U_1 - \mathbb{E}\{U_1 | U_0\} = U_1 - \Delta\Gamma^{-1} U_0$$

and $V_1 \sim N_m(0, \Psi)$, where $\Psi = \bar\Omega - \Delta\Gamma^{-1}\Delta^\top$. The dual transformation

$$U_0 = V_0, \qquad U_1 = \Delta\Gamma^{-1}V_0 + V_1$$

achieves the required equality of the correlation matrices provided $\bar\Omega = \Delta\Gamma^{-1}\Delta^\top + \Psi$. In other words, we are taking

$$A_0 = I_m, \quad A_1 = 0, \quad B_0 = \Delta\Gamma^{-1}, \quad B_1 = I_d.$$

in (15).

We now use these relationships to establish an explicit connection between the two stochastic representations of the SUN indicated in Section 2.1. Consider (10) and notice that the distribution of $V_0(-\gamma)$ is the same of $(V_0|V_0 + \gamma > 0)$, if $V_0 \sim N_m(0, \Gamma)$. Now re-write the convolution form in (11) as

$$\begin{aligned} B_0V_0(-\gamma) + B_1V_1 &= (B_0V_0 + B_1V_1|V_0 + \gamma > 0) \\ &= (B_0U_0 + B_1(U_1 - \Delta\Gamma^{-1}U_0)|U_0 + \gamma > 0) \\ &= (U_1|U_0 + \gamma > 0) \end{aligned} \tag{16}$$

under the above specifications of B_0, B_1.

This argument shows that the representation via convolution for V variables corresponds to a representation via conditioning for the U variables. Also, by reading the sequence of equalities backward, we transform the representation via conditioning into the one via convolution.

An analogous analysis can be carried out for the related variants of SN distribution. The following table refers to those variants introduced via a conditioning mechanism operating, and it provides the ingredients of (15) that relate the U components used for conditioning to underlying independent variables V_0, V_1.

For each of these families, it is possible to convert the representation via conditioning into one of convolution type, by introducing orthogonal components $V_0 = U_0$, $V_1 = U_1 - \mathbb{E}\{U_1|U_0\}$, similarly to the SUN family and then replicating the argument in (16).

The CSN-2 case leads to a particularly clear-cut form, as follows. Let $U_0 = V_0$, $U_1 = U_1 - \mathbb{E}\{U_1|U_0\} = U_1 - D^\top U_0$, hence $U_1 = D^\top V_0 + V_1$. On making use of the moment generating function, it can be easily be shown that

$$Z = (U_1|U_0 + \gamma > 0) = (D^\top V_0 + V_1|V_0 + \gamma > 0) = D^\top(V_0|V_0 + \gamma) + V_1 = D^\top V_0(-\gamma) + V_1,$$

and obtain that $Z = D^\top V_0(-\gamma) + V_1 \sim \mathrm{CSNII}_{d,m}(0, \gamma, \Psi, \Gamma)$. This produces a representation of the CSN-2 family of convolution type, with a natural choice of the matrices B_0 and B_1 in (12).

The table below summarises the expressions of the matrices involved by the above correspondence for the various families.

	A_0	A_1	B_0	B_1	$\mathrm{var}\{V_0\}$	$\mathrm{var}\{V_1\}$
CSN-1	I_m	0	$\Sigma D^\top(\Theta + D\Sigma D^\top)^{-1}$	I_d	$\Theta + D\Sigma D^\top$	$(\Sigma^{-1} + D\Theta^{-1}D^\top)^{-1}$
CSN-2	I_m	0	$D^\top$	I_d	Γ	Ψ
HSN	I_m	0	$-\Upsilon C^\top(C\Upsilon C^\top)^{-1}$	I_d	$C\Upsilon C^\top$	$\Sigma + \Upsilon - \Upsilon C^\top(C\Upsilon C^\top)^{-1}C\Upsilon$

2.4 Other properties of the SUN

We do not discuss in detail formal properties of the SUN family, as these coincide with those of the related families discussed earlier, once the parameterisations are properly selected.

Moments and cumulants can be obtained directly from (13), or from suitable adaption of expressions given by Gupta et al. (2004). The derivation of moments is simplified when $\Gamma = \mathrm{diag}(\tau_1^2, \ldots, \tau_m^2)$, since the cumulants generating function reduces to

$$K(t) = \log M(t) = \xi^\top t + \tfrac{1}{2} t^\top \Omega t + \sum_{j=1}^{m} \log \Phi(\tau_j^{-1}\gamma_j + \tau_j^{-1}\delta_{\cdot j}^\top \omega t) - \log \Phi(\gamma; \Gamma).$$

where $\delta_{\cdot 1}, \ldots, \delta_{\cdot m}$ are the columns of Δ. From this expression, we obtain

$$\mathbb{E}\{Y\} = K'(0) = \xi + \sum_{j=1}^{m} \zeta_1(\tau_j^{-1}\gamma_j)\, \tau_j^{-1}\omega\delta_{\cdot j}$$

and

$$\mathrm{var}\{Y\} = K''(0) = \Omega + \sum_{j=1}^{m} \zeta_2(\tau_j^{-1}\gamma_j)\, \tau_j^{-2}\omega\delta_{\cdot j}\delta_{\cdot j}^\top \omega$$

where $\zeta_r(x)$ is the r-th derivative of $\zeta_0(x) = \log\{2\,\Phi(x)\}$.

The distribution of a quadratic form of type $Q(Z) = Z^\top A Z$, where $Z \sim \mathrm{SUN}_{d,m}(0, \gamma, 1_d, \Omega^*)$ and A is a $d \times d$ symmetric matrix of rank p. It can be shown that the moment generating function of $Q(Z)$ is

$$M_Q(t) = |I_d - 2\,t\,A\bar{\Omega}|^{-1/2} \frac{\Phi_m(\gamma; \Gamma + 2t\Delta^\top (I_d - 2tA\bar{\Omega})^{-1} A\Delta)}{\Phi_m(\gamma; \Gamma)}.$$

An important special case is when $A = \bar{\Omega}^{-1}$. To obtain $Q(Z) \sim \chi_d^2$ similarly to the case when Z has normal or skew-normal distribution, we need the conditions $\gamma = 0$ and that Γ and $\Delta^\top \bar{\Omega}^{-1} \Delta$ are diagonal matrices. However, since $|I_d - 2tA\bar{\Omega}|^{-1/2} = (1-2t)^{-1/2} \Leftrightarrow A\bar{\Omega}A = A$, it follows that $Q(Z) \sim \chi_p^2 \Leftrightarrow A\bar{\Omega}A = A$ and $A\Delta = 0$.

3 Extension to the skew-elliptical family

The purpose of the present section is extend the results of the previous sections from the normal framework to the one of elliptical distributions, at least to some extent. In the development, we shall make use of some standard results on elliptical distributions, which we do not reproduce here in detail, as they can be found in standard references; see for instance Fang et al. (1990).

To start with, replace the normality assumption in (7) by the assumption of elliptical density; hence write

$$U = \begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim El_{m+d}(0, \Omega^*; h_{m+d}), \qquad \Omega^* = \begin{pmatrix} \Gamma & \Delta^\top \\ \Delta & \bar{\Omega} \end{pmatrix},$$

where again Ω^* is a correlation matrix. Here h_{m+d} is the so-called density generator, such that the density of U is

$$f_{m+d}(u;\Omega^*) = |\Omega^*|^{-1/2}\, h_{m+d}\left(u^\top(\Omega^*)^{-1}u\right), \tag{17}$$

for $u \in \mathbb{R}^{m+d}$; we shall denote its distribution function by $F_{m+d}(u;\Omega^*)$. The two components of U have distribution of type

$$U_0 \sim El_m(0,\Gamma;h_m), \qquad U_1 \sim El_m(0,\bar\Omega;h_d),$$

whose density generators h_m and h_d are computed from h_{m+d} using formula (2.23) of Fang et al. (1990); replacing these generators in (17) in place of h_{m+d} provides the expressions of their densities. The conditional distribution of U_0 given $U_1 = y$ is

$$(U_0|U_1 = y) = (U_0|q(U_1) = q(y)) \sim EL_m(\Delta^\top\bar\Omega^{-1}y, \Gamma - \Delta^\top\bar\Omega^{-1}\Delta, h_{m;q(y)}),$$

where $q(y) = y^\top\bar\Omega^{-1}y$ and $h_{m;a}(u) = h_{m+d}(u+a)/h_d(a)$ denotes the density generator.

Proceeding in the same fashion of the normal case, we consider the variable $Y = \xi + \omega Z$, where $Z = (U_1|U_0 + \gamma > 0)$, and use the notation $Y \sim \mathrm{SUEl}_{d,m}(\xi,\gamma,\bar\omega,\Omega^*;h_{m+d})$. Its density function can be computed inserting the above ingredients in (8), leading to

$$f(y) = f_d(y-\xi;\Omega)\frac{F_{m;q[\omega^{-1}(y-\xi)]}(\gamma + \Delta^\top\bar\Omega^{-1}\omega^{-1}(y-\xi);\Gamma - \Delta^\top\bar\Omega^{-1}\Delta)}{F_m(\gamma;\Gamma)}, \tag{18}$$

where $F_{m;a}(x;\Theta)$ denotes the distribution function of the $El_m(0,\Theta,h_{m;a})$ distribution, namely

$$F_{m;a}(x;\Theta) = |\Theta|^{-1/2}\int_{v\le x}\frac{h_{m+d}(v^\top\Theta^{-1}v + a)}{h_d(a)}\,\mathrm{d}v.$$

Various variant forms can now be obtained from (18) similarly to the normal context above. We focus attention on the restriction $\gamma = 0$, which has played an important role also for the normality case. Under this assumption, the simplification $F_m(0;\Gamma) = \Phi_m(0;\Gamma)$ holds. This fact can easily be established taking into account the representation of an elliptical variate X as $X = RW$ where R denotes a positive scalar variable whose distribution depends of the density generator of X, and W is uniformly distributed on the unit sphere in $\mathbb{R}^k$, if k is the dimension of X. It is immediately seen that the probability assigned to each cone with vertex in the origin is the same for all possible distributions of R, and this property holds in particular for the orthant; since this probability does not depend on the generator, we can compute it assuming normality.

An important class of elliptical distributions is generated by the scale-mixtures of the normal variables. For any elliptical distribution in this class, there is a non-negative random variable S such that $U|S = s \sim N_{m+d}(0, s^{-1}\Omega^*)$, with S having distribution function H. In this case

$$(Z|S = s) = (U_1|U_0 + \gamma > 0, S = s) \sim \mathrm{SUN}_{d,m}(0,\gamma,1_d,s^{-1}\Omega^*),$$

and, using (9), the unconditional density of Z is

$$g(z) = \int_0^\infty s^{d/2}\phi_d(\sqrt{s}\,z;\bar\Omega)\,\frac{\Phi_m(\sqrt{s}\,\gamma + \sqrt{s}\,\Delta^\top\bar\Omega^{-1}z;\Gamma - \Delta^\top\bar\Omega^{-1}\Delta)}{\Phi_m(\sqrt{s}\,\gamma;\Gamma)}\,\mathrm{d}H(s),$$

which for $\gamma = 0$ reduces to

$$g(z) = \frac{1}{\Phi_m(0;\Gamma)} \int_{x\leq \Delta^\top\bar\Omega^{-1}z} \int_0^\infty s^{(m+d)/2}\phi_d(\sqrt{s}\,z;\bar\Omega)\,\phi_m(\sqrt{s}\,x;\Gamma-\Delta^\top\bar\Omega^{-1}\Delta)\,\mathrm{d}H(s)\,\mathrm{d}x\,,$$

from which one can obtain the density of $Y = \xi + \omega Z$, at least for some convenient forms of $H(s)$.

An important special case of the above density $g(z)$ occurs when $S \sim Gamma(\nu/2, \nu/2)$, since it is well-known that this mixing distribution applied to the normal variables produces the multivariate t distribution. After some algebraic manipulations, similar to those employed by Arellano-Valle & Bolfarine (1995), in this case $g(z)$ becomes

$$\begin{aligned} g_T(z) &= \frac{c_{m+d,\nu}}{\Phi_m(0;\Gamma)\,|\bar\Omega|^{1/2}\,|\Gamma-\Delta^\top\bar\Omega^{-1}\,\Delta|^{1/2}} \\ &\qquad \times \int_{x\leq\Delta^\top\bar\Omega^{-1}z} \{\nu+q(z)+x^\top(\Gamma-\Delta^\top\bar\Omega^{-1}\Delta)^{-1}x\}^{-(m+d+\nu)/2}\,\mathrm{d}x \\ &= \frac{1}{\Phi_m(0;\Gamma)}\,t_d(z;\bar\Omega,\nu)\int_{x\leq\Delta^\top\bar\Omega^{-1}z} t_m\left(x;\frac{\nu+q(z)}{\nu+d}(\Gamma-\Delta^\top\bar\Omega^{-1}\Delta),\nu+d\right)\,\mathrm{d}x \\ &= \frac{1}{\Phi_m(0;\Gamma)}\,t_d(z;\bar\Omega,\nu)\,T_m\left(\Delta^\top\bar\Omega^{-1}z;\frac{\nu+q(z)}{\nu+d}(\Gamma-\Delta^\top\bar\Omega^{-1}\Delta),\nu+d\right) \end{aligned}$$

where $q(z)$ is defined above,

$$c_{k,\upsilon} = \frac{\Gamma[\frac{1}{2}(k+\upsilon)]\,\upsilon^{\upsilon/2}}{\Gamma(\frac{1}{2}\upsilon)\,\pi^{k/2}},$$

and $t_k(x;\Theta,\upsilon)$, $T_k(x;\Theta,\upsilon)$ denote the density function and the distribution function, respectively, of a k-dimensional t-distribution with dispersion matrix Θ and υ degrees of freedom.

We have then obtained a d-dimensional density function of skew-t type. When $m = 1$, this $g_T(z)$ reduces to a form of skew-t already proposed in the literature. This connection is more directly visible if one compares the density $g_T(z)$ with formula (26) of Azzalini & Capitanio (2003) and formula (2.3) of Gupta (2003); these densities are in turn equivalent to the skew-t density of Branco & Dey (2001), although expressed in a different form.

Under the assumption $\gamma = 0$, it is possible to establish a connection between the representation of $\mathrm{SUEl}_{d,m}$ distribution via conditioning examined above and a representation via convolution. Assume

$$\begin{pmatrix} V_0 \\ V_1 \end{pmatrix} \sim El_{m+d}(0,\Sigma;h_{m+d}), \quad \Sigma = \begin{pmatrix} \Gamma & 0 \\ 0 & \Psi \end{pmatrix},$$

and, similarly the normal case, consider $V_0 = U_0$ and $V_1 = U_1 - \mathbb{E}\{U_1|U_0\}$, i.e.

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} = \begin{pmatrix} V_0 \\ \Delta\Gamma^{-1}V_0 + V_1 \end{pmatrix} \sim El_{m+d}(0,\Omega^*;h), \quad \Omega^* = \begin{pmatrix} \Gamma & \Delta^\top \\ \Delta & \bar\Omega \end{pmatrix},$$

where $\bar\Omega = \Delta\Gamma^{-1}\Delta^\top + \Psi$. Note that, except in the case when h_{m+d} is the normal generator, the variables V_0 and V_1 are uncorrelated but not independent.

In fact, define $V_0(c) = (V_0|V_0 + c > 0) \sim LTEl_m(c; 0, \Gamma; h_m)$, where $V_0 \sim El_m(0, \Gamma; h_m)$ and consider the well-known stochastic representation, stated for instance in Theorem 2.6 of

Fang et al. (1990),

$$\begin{pmatrix} V_0 \\ V_1 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \Gamma^{1/2} R T_0 W_0 \\ \Psi^{1/2} R T_1 W_1 \end{pmatrix},$$

where R, T_0, W_0 and W_1 are independent variables, such that the so-called radial variable R is scalar positive, $T_0^2 \sim Beta(m/2, m/2)$ and W_0 and W_1 have uniform distribution on the unit sphere of dimension m and d, respectively; $T_1 = \sqrt{1 - T_0^2}$. Thus, it follows that

$$\begin{aligned} (B_0 V_0 + B_1 V_1 | V_0 > 0) &\stackrel{d}{=} (B_0 \Gamma^{1/2} T_0 W_0 + B_1 \Psi^{1/2} R T_1 W_1 | \Gamma^{1/2} R T_0 W_0 > 0) \\ &= (B_0 \Gamma^{1/2} R T_0 W_0 + B_1 \Psi^{1/2} R T_1 W_1 | \Gamma^{1/2} W_0 > 0) \\ &\stackrel{d}{=} B_0 \Gamma^{1/2} R T_0 (W_0 | \Gamma^{1/2} R T_0 W_0 > 0) + B_1 \Psi^{1/2} R T_1 W_1 \\ &\stackrel{d}{=} B_0 V_0(0) + B_1 V_1. \end{aligned}$$

Thus, $Z' = B_0 V_0(0) + B_1 V_1 \sim SUEl_{d,m}(0, 0, 1_d, \Omega^*)$, provided conditions (12) are satisfies. For $B_1 = I_d$, we have Z' and $Z = (U_1 | U_0 > 0)$ have the same SUEl if $\bar{\Omega} = \Delta \Gamma^{-1} \Delta^\top + \Psi$.

Acknowledgements

A substantial part of this paper has been developed while the second author was visiting the Departamento de Estadística de la Pontificia Universidad Católica de Chile, whose kind hospitality is gratefully acknowledged. The work of the first author has been supported by grant Fondecyt No. 1040865.

References

Arellano-Valle, R. B. & Bolfarine, H. (1995). On some characterizations of the t-distribution. *Statist. & Prob. Lett.* 25, 79–85.

Arellano-Valle, R. B., del Pino, G. & San Martín, E. (2002). Definition and probabilistic properties of skew-distribution. *Statist. & Prob. Lett.* 58, 111–121.

Arellano-Valle, R. B. & Genton, M. G. (2005). On fundamental skew distributions. To appear in *Journal of Multivariate Analysis.*

Arnold, B.C. and Beaver, R.J. (2000). Hidden truncation models. *Sankhya* 62, 23-25.

Arnold, B.C. and Beaver, R.J. (2002). Skewed multivariate models related to hidden truncation. *Test* 11, 7-54.

Arnold & Beaver (2002). Alternative construction of skewed multivariate distributions. Unpublished.

Azzalini, A. (2003). Skew-normal family of distributions. *Encyclopedia of Statistical Sciences*, 2nd edition. J. Wiley & Sons, New York. To appear.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distributions. *J. R. Statist. Soc. B* **61**, 579–602.

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–26.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew elliptical distributions. *Journal of Multivariate Analysis* **79**, 99–113.

González-Farías, G., Domínguez-Molina, A. & Gupta, A. K. (2004). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference* **126**, 521–534..

Gupta, A. K. (2003). Multivariate skew t-distribution. *Statistics* 37, 359–363.

Gupta, A. K., González-Farías, G., & Domínguez-Molina, A. (2004). A multivariate skew normal distribution, *Journal of Multivariate Analysis* 89, 181–190.

Liseo, B. & Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. *Statistics & Probability Letters* 61, 395–401.

Nelson, Lloyd S. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics* **6**, 469–471.

Sahu, S. K., Dey, D. K. & Branco, M. D. (2003) A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian J. Statistics* 31, 129–150.

Weinstein, M. A. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics* **6**, 104–105.

Appendix A: Conditions for independence

We examine conditions under which a partition of a random variable $\text{SUN}_d(0, \gamma, 1_d, \Omega^*)$ into two components attains independence of these two variables. The question is elementary if we allow either of these variables to be regular normal, since the conditions are the same for independence between multivariate normal variables. We then discuss the case where both subcomponents are allowed to be non-symmetric.

For (U_0, U_1) distributed as in (7), set $Z = (U_1 | U_0 + \gamma > 0)$, and consider the partition of Z in two components of size d_1 and d_2, respectively, with $d_1 + d_2 = d$. Correspondingly, write

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \quad U_1 = \begin{pmatrix} U_{11} \\ U_{12} \end{pmatrix}, \quad \Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}, \quad \bar{\Omega} = \begin{pmatrix} \bar{\Omega}_{11} & \bar{\Omega}_{12} \\ \bar{\Omega}_{21} & \bar{\Omega}_{22} \end{pmatrix};$$

hence $Z_i = (U_{1i} | U_0 + \gamma > 0)$, or equivalently $Z_i = A_i Z = (A_i U_1 | U_0 + \gamma > 0)$, for $i = 1, 2$, where $A_1 = (I_{d_1}, 0)$, $A_2 = (0, I_{d_2})$. Clearly $Z_i \sim \text{SUN}_{d_i,m}(0, \gamma, 1_{d_i}, \Omega_i^*)$,

$$\Omega_i^* = \begin{pmatrix} \Gamma & \Delta^\top A_i^\top \\ A_i \Delta & A_i \bar{\Omega} A_i^\top \end{pmatrix} = \begin{pmatrix} \Gamma & \Delta_i^\top \\ \Delta_i & \bar{\Omega}_{ii} \end{pmatrix}, \qquad (i = 1, 2).$$

Consider a similar partition of U_0 in blocks of size m_1 and m_2, respectively, $(m_1 + m_2 = m)$ and write

$$U_0 = \begin{pmatrix} U_{01} \\ U_{02} \end{pmatrix}, \quad \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}, \quad \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix},$$

where U_{0i} and γ_i have dimension $m_i \times 1$, Γ_{ij} has dimension $m_i \times m_j$, and Δ_{ij} has dimension $d_i \times m_j$ $(i, j = 1, 2)$.

Since

$$Z_1 = (U_{11} | U_{01} + \gamma_1 > 0, U_{02} + \gamma_2 > 0), \qquad Z_2 = (U_{12} | U_{01} + \gamma_1 > 0, U_{02} + \gamma_2 > 0),$$

then there are two situations under which Z_1 and Z_2 are independent:

(i) if (U_{11}, U_{01}) and (U_{12}, U_{02}) are independent, i.e.

$$Z_1 \stackrel{d}{=} (U_{11} | U_{01} + \gamma_1 > 0), \quad Z_2 \stackrel{d}{=} (U_{12} | U_{02} + \gamma_2 > 0),$$

which occurs if

$$\Gamma_{12} = 0, \quad \bar{\Omega}_{12} = 0, \quad \Delta_{12} = 0, \quad \Delta_{21} = 0;$$

(ii) if (U_{11}, U_{02}) and (U_{12}, U_{01}) are independent, i.e.

$$Z_1 \stackrel{d}{=} (U_{11} | U_{02} + \gamma_2 > 0), \quad Z_2 \stackrel{d}{=} (U_{12} | U_{01} + \gamma_1 > 0),$$

which occurs if

$$\Gamma_{12} = 0, \quad \bar{\Omega}_{12} = 0, \quad \Delta_{11} = 0, \quad \Delta_{22} = 0.$$

The above conditions are sufficient for independence; we now turn to necessary conditions. It is clear that the conditions $\bar{\Omega}_{12} = \mathrm{cov}\{U_{11}, U_{12}\} = 0$, $\Gamma_{12} = \mathrm{cov}\{U_{01}, U_{02}\} = 0$ are necessary, since we need U_{11} to be independent of U_{12} and U_{01} independent of U_{02}, respectively, in order that $\phi_d(z; \bar{\Omega})$ factorises as $\phi_{d_1}(z_1; \bar{\Omega}_{11})\, \phi_{d_2}(z_2; \bar{\Omega}_{22})$ and $\Phi_m(\gamma; \Gamma)$ factorises as $\Phi_{m_1}(\gamma_1; \Gamma_{11})\, \Phi_{m_2}(\gamma_2; \Gamma_{22})$.

In order to examine the conditions on the matrix Δ, let

$$\bar{\Omega}^{-1} = \begin{pmatrix} \bar{\Omega}^{11} & \bar{\Omega}^{12} \\ \bar{\Omega}^{21} & \bar{\Omega}^{22} \end{pmatrix},$$

and, under the conditions $\bar{\Omega}_{12} = 0$, $\Gamma_{12} = 0$, we have that

$$\gamma + \Delta^\top \bar{\Omega}^{-1} z = \begin{pmatrix} \gamma_1 + \Delta_{11}^\top \bar{\Omega}_{11}^{-1} z_1 + \Delta_{21}^\top \bar{\Omega}_{22}^{-1} z_2 \\ \gamma_2 + \Delta_{12}^\top \bar{\Omega}_{11}^{-1} z_1 + \Delta_{22}^\top \bar{\Omega}_{22}^{-1} z_2 \end{pmatrix},$$

and

$$\Gamma - \Delta^\top \bar{\Omega}^{-1} \Delta = \begin{pmatrix} \Gamma_{11} - (\Delta_{11}^\top \bar{\Omega}_{11}^{-1} \Delta_{11} + \Delta_{21}^\top \bar{\Omega}_{22}^{-1} \Delta_{21}) & -(\Delta_{11}^\top \bar{\Omega}_{11}^{-1} \Delta_{12} + \Delta_{21}^\top \bar{\Omega}_{22}^{-1} \Delta_{22}) \\ -(\Delta_{12}^\top \bar{\Omega}_{11}^{-1} \Delta_{11} + \Delta_{22}^\top \bar{\Omega}_{22}^{-1} \Delta_{21}) & \Gamma_{22} - (\Delta_{12}^\top \bar{\Omega}_{11}^{-1} \Delta_{12} + \Delta_{22}^\top \bar{\Omega}_{22}^{-1} \Delta_{22}) \end{pmatrix}.$$

Therefore, $\Phi_m(\gamma + \Delta^\top \bar{\Omega}^{-1} z; \Gamma - \Delta^\top \bar{\Omega}^{-1} \Delta)$ can be factorised in either of the following two ways:

(i) when $\Delta_{21}^\top \bar{\Omega}_{22}^{-1} = 0$ and $\Delta_{12}^\top \bar{\Omega}_{11}^{-1} = 0$, as

$$\Phi_{m_1}(\gamma_1 + \Delta_{11}^\top \bar{\Omega}_{11}^{-1} z_1; \Gamma_{11} - \Delta_{11}^\top \bar{\Omega}_{11}^{-1} \Delta_{11}) \Phi_{m_2}(\gamma_2 + \Delta_{22}^\top \bar{\Omega}_{22}^{-1} z_2; \Gamma_{22} - \Delta_{22}^\top \bar{\Omega}_{22}^{-1} \Delta_{22});$$

(ii) or when $\Delta_{11}^\top \bar\Omega_{11}^{-1} = 0$ and $\Delta_{22}^\top \bar\Omega_{22}^{-1} = 0$, as

$$\Phi_{m_1}(\gamma_1 + \Delta_{21}^\top \bar\Omega_{22}^{-1} z_1; \Gamma_{11} - \Delta_{21}^\top \bar\Omega_{22}^{-1}\Delta_{21})\Phi_{m_2}(\gamma_2 + \Delta_{12}^\top \bar\Omega_{11}^{-1} z_2; \Gamma_{22} - \Delta_{12}^\top \bar\Omega_{11}^{-1}\Delta_{12}).$$

Consequently, the conditions for independence are either of the following set:

(i) $\bar\Omega_{12} = 0$, $\Gamma_{12} = 0$, $\Delta_{21}^\top \bar\Omega_{22}^{-1} = 0$ and $\Delta_{12}^\top \bar\Omega_{11}^{-1} = 0$;

(ii) $\bar\Omega_{12} = 0$, $\Gamma_{12} = 0$, $\Delta_{11}^\top \bar\Omega_{11}^{-1} = 0$ and $\Delta_{22}^\top \bar\Omega_{22}^{-1} = 0$.

Notice that, for the above conditions refer to a given partition of Z and a given partition of U_0. However, the actual question of independence between Z_1 and Z_2 is not linked to a specific partition of U_0. Since, if $m > 2$, the partition of U_0 in two subcomponents is not unique, the above conditions must be checked for all possible partitions of U_0. Luckily m can be expected to be very small in most cases, as otherwise the mere computation of the density (9) becomes unfeasible; hence the number of cases to be scanned is limited.